Folded-back solution structure of monomeric Factor H of human complement by scattering, ultracentrifugation and modelling

Mohammed Aslam and Stephen J. Perkins

Department of Biochemistry and Molecular Biology, Royal Free and University College Medical School, University College London, Gower Street, London WC1E 6BT, U.K.

Received 19th February 2002; accepted 16th April 2002.

Introduction

Factor H (FH) of the complement system of immune defence is present in plasma at about 0.5 mg/ml. It consists entirely of 20 short complement/consensus repeat (SCR) domains, each of length about 61 residues, where SCR domains constitute the most abundant domain type in the complement proteins. The principal function of FH is to regulate the alternative pathway of complement activation by acting as a cofactor for factor I in the breakdown of C3b to form iC3b. The cofactor and decay accelerating activity are located within the four N-terminal domains, SCR-1 to SCR-4, which bind to intact C3b. A second C3 site is located within SCR-6 to SCR-10 which binds to the C3c region of C3b, and a third site is located within SCR-16 and SCR-20 which binds to the C3d region of C3b. FH also binds to heparin and other polyanionic substrates, where heparin modulates the complement regulatory functions of FH. Two heparin binding sites have been located in SCR-7 and SCR-20 in recombinant FH, and a third heparin binding site is suggested to be located at or near SCR-13. The SCR domains are thought to act synergistically to enable FH to achieve differential control of complement activation.

Molecular structures of FH are required to clarify its mechanism of action. However, its crystallisation is thought to be precluded by the 19 potential flexible peptide links between the 20 SCR domains and the existence of up to 9 putative glycosylation sites. Recent NMR and crystal structures for other proteins with two to five SCR domains showed that there is no preferred structural orientation between two adjacent SCR domains, thus it is not possible to deduce an overall structure for FH from these studies. Accordingly, to determine a medium-resolution molecular structure for FH, we have used x-ray and neutron scattering in conjunction with automated molecular modelling methods to assess these interdomain linkers. The FH curve fits were highly constrained by the use of 17 homology models and three known NMR structures for the 20 SCR domains of FH [1]. The present article reviews the scattering approach that was used for FH in terms of our earlier applications of these automated rigid-body curve fit searches. The novelty of this strategy is that it is applied here for the first time to a linear arrangement of 20 domains [see refs 2-4]. Biologically useful information can be extracted from the medium-resolution structures obtained by scattering analyses [5].

X-ray and neutron scattering and ultracentrifugation of FH

X-ray and neutron scattering was performed with FH in the concentration range between 0.7-14 mg/ml. The joint use of x-rays and neutrons provided independent checks of self-consistency in the scattering curves. It was already known that FH is highly elongated in solution [6], thus the camera at Station 2.1 at SRS Daresbury was set up with sample-detector distances of up to 5.64 m in order to ensure that a sufficiently low Q had been obtained for Guinier analyses. For the same reason, after initial data collection on the LOQ camera at ISIS had verified its monomeric state, the final neutron scattering data were obtained on Instrument D11 with a sample-detector distance of 10 m and Instrument D22 with a sample-detector distance of 8 m, both at the Institut-Laue-Langevin. Full details are given in [1]. From the x-ray and neutron Guinier plots, the radius of gyration R_G of FH was determined to be 11.1-11.3 nm. These similar values indicated that radiation damage effects sometimes seen by x-rays were absent. Cross-sectional Guinier analyses revealed two linear regions. From these, two distinct radii of gyration of the cross-section R_{XS}-

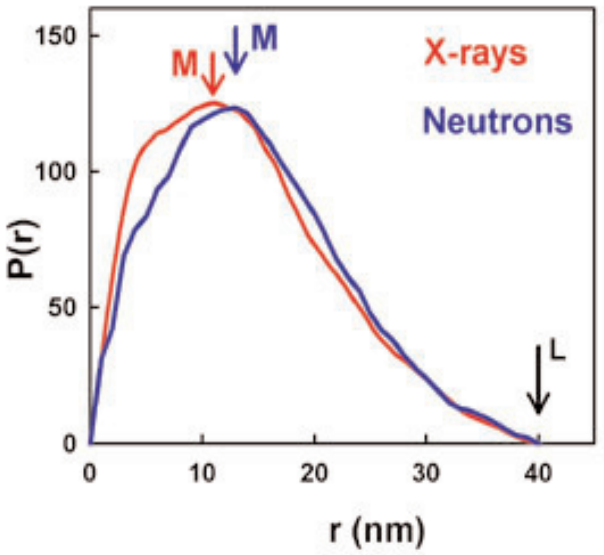

Figure 1: Experimental X-ray and neutron distance distribution functions P(r) for FH.

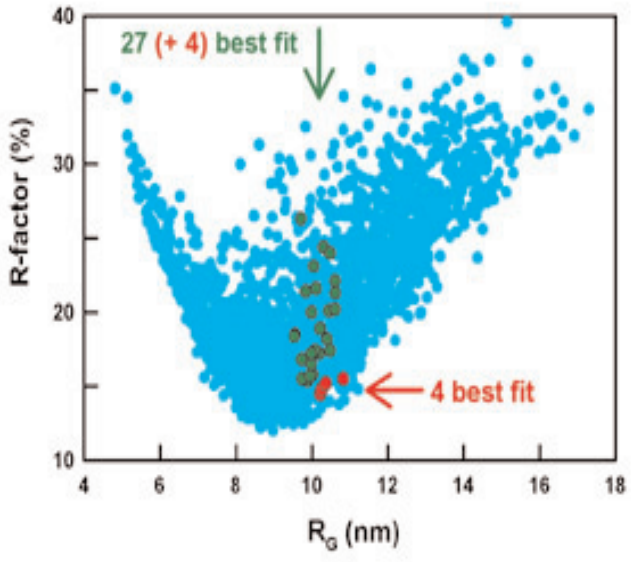

Figure 2: Comparison of the goodness-of-fit R-factor values for 2010 FH models with their RG values in one of the eight full automated conformational searches for a FH solution structure. The best fit solutions are highlighted in green and red.

$_1$ and R_{XS-2} were identified in two Q ranges to be 4.4 nm and 1.7 nm respectively. These were assigned to an averaged higher-order structure formed by the proximity of non-adjacent SCR domains in FH (R_{XS} of 4.4 nm) and the averaged cross-sectional structure of a line of SCR domains (R_{XS} of 1.7 nm) (see below). The x-ray and neutron distance distribution functions *P(r)* for FH were of particular interest. In these, the most frequently occurring distance within FH (M) are similar at 11 nm (x-rays) and 12 nm (neutrons), and the maximum dimension is denoted by *L* which is 40 nm for both the x-ray and neutron data (Figure 1). If all the SCR domains in FH were fully extended in conformation, this would result in a length *L* of 73 nm for FH. This difference in the experimental and predicted *L* values shows that the 20 SCR domains of FH are folded back upon themselves in solution. Other analyses based on the neutron I(0)/c data from Guinier plots and from sedimentation equilibrium fits by analytical ultracentrifugation showed that FH was monomeric with a molecular weight of 165,000 ± 17,000 (neutrons) and 148,000-159,000 (ultracentrifugation). The use of sedimentation velocity experiments showed that FH had a sedimentation coefficient of 5.3 ± 0.1 S. The combination of this value of 5.3 S with the RG determination of 11.1-11.3 nm in the modelling below provided an independent confirmation that the correct Guinier Q range had been identified for the RG determination. This was necessary as there was a chance that the true Guinier region for FH might have occurred at even lower Q values than that utilised here.

Automated modelling of the FH structure

Advances in modelling calculations have shown that scattering curves and analytical ultracentrifugation data are fully calculable from atomic structures [2,3]. Most notably, this method has been applied to determine the relative location of the two Fab and Fc fragments within antibody structures. Each of the Fab and Fc fragments was treated as a rigid body of known atomic structure. The modelling analysis identified the conformation of the two hinge peptides that connected each Fab-Fc pair by evaluating 12,000 variable conformations for these [4]. In comparison to antibody modelling, the major difference in the modelling of FH is the need to incorporate 19 variable linker peptide conformations to link the 20 SCR domains, not one. Despite the multiparameter nature of this new analysis, it was nonetheless possible to determine molecular structures from the scattering and ultracentrifugation data.

In order to initiate the FH modelling for the scattering curve and sedimentation coefficient fits, homology models were constructed for 17 of the 20 SCR domains using known NMR structures for FH SCR-5, SCR-15 and SCR-16 and other homologues. Trial calculations then evaluated the modelling strategy using different types of links. One FH model

was created using extended β-strands to connect the SCR domains. The known linker conformations from NMR and crystallography were used to make seven other FH models. While none of these models gave good scattering curve fits, they provided important clues. In particular, the RXS-1 value was found to be sensitive to the higher-order SCR structure while the RXS-2 value was not sensitive to the linker conformation provided that this was extended. In order to perform an automated search to identify best-fit FH structures, molecular dynamics simulations were used to generate libraries of 10,000 random conformations for each of the 19 SCR-SCR linker peptides. Peptides from these libraries were randomly selected for combination with the 20 SCR structures in order to generate stereochemically-complete models for the FH structure. In the automated curve fit procedure, a total of 16,752 FH models were generated from eight different searches of about 2000 models each, in each of which the assumptions made for the linker conformations were varied.

The typical outcome of one of the eight searches is summarised for 2010 FH models in Figure 2 in which a minimum length was specified for each of the 19 linkers (see [1]). Filters based on the experimental values for the R_G, R_{XS-1} and R_{XS-2} values from scattering and the sedimentation coefficient from analytical ultracentrifugation were used to retain or reject candidate FH models. Few models survived these filters, which illustrated the computational difficulty in identifying a good scattering fit model. The 31 best-fit FH models all had RG values close to 11 nm (green and red in Figure 2). The lengths of these models ranged between 8 nm to 33 nm (see [1]), which is far less than the distance of 73 nm predicted for the most extended FH structure possible. From these 31 models, the four best-fit FH ones (red in Figure 2) were identified by their lowest *R*-factor values of between 12% to 15% in Figure 2b. Since the *R*-factors are generally between 1.2-8.7% for good scattering fit models [2], this showed that the four best-fit FH models were limited in their ability to describe the X-ray and neutron curves. Nonetheless it is concluded that only those FH models in which the 20 SCR domains were bent back upon themselves will fit the scattering and ultracentrifugation data.

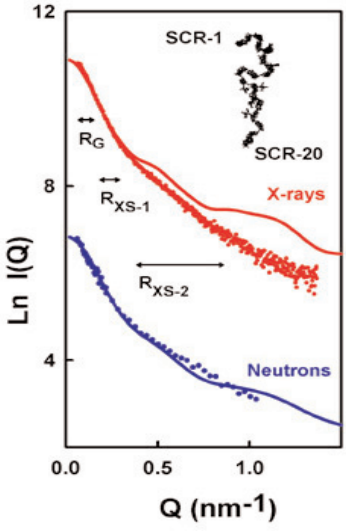

Figure 3: X-ray and neutron curve fits for one of the four best-fit models for FH that was identified from the automated search of FH models summarised in Figure 2.

Typical x-ray and neutron curve fits for one of the four best-fit FH models are shown in Figure 3, together with the α-carbon trace of the FH model in question. The upper curve fit corresponds to x-ray data from Station 2.1 at the SRS Daresbury, while the lower curve fit corresponds to neutron data from Instrument D11 at the Institut-Laue-Langevin. The continuous lines correspond to the calculated curve from each model. The arrowed ranges indicate the *Q* ranges used for the R_G, R_{XS-1} and R_{XS-2} analyses. These show that the overall structure as monitored experimentally by the R_G and R_{XS-1} parameters is well modelled, while the transition point in the scattering curve between the R_{XS-1} and R_{XS-2} regions is less well modelled. One explanation of this discrepancy is the inability to generate a sufficiently wide range of FH conformers in the computer time available for this project. It is also possible that, if there is sufficient interdomain flexibility in FH in solution, this would reduce both the detail in the FH scattering curve at large *Q* and the applicability of static models to represent this region of the curve. The four best-fit FH models are shown in Figure 4, in which the SCR-1 domains (grey) of the four models are superimposed. This indicates the folded back conformation of the FH structures, together with the random orientations of SCR-2 to SCR-20 that were generated in these models.

Conclusions

The solution data and their modelling show that the SCR domains in FH are folded back. These bent-back domain structures may correspond to conformational flexibility in FH that may enable the multiple FH binding sites for C3 and heparin to come

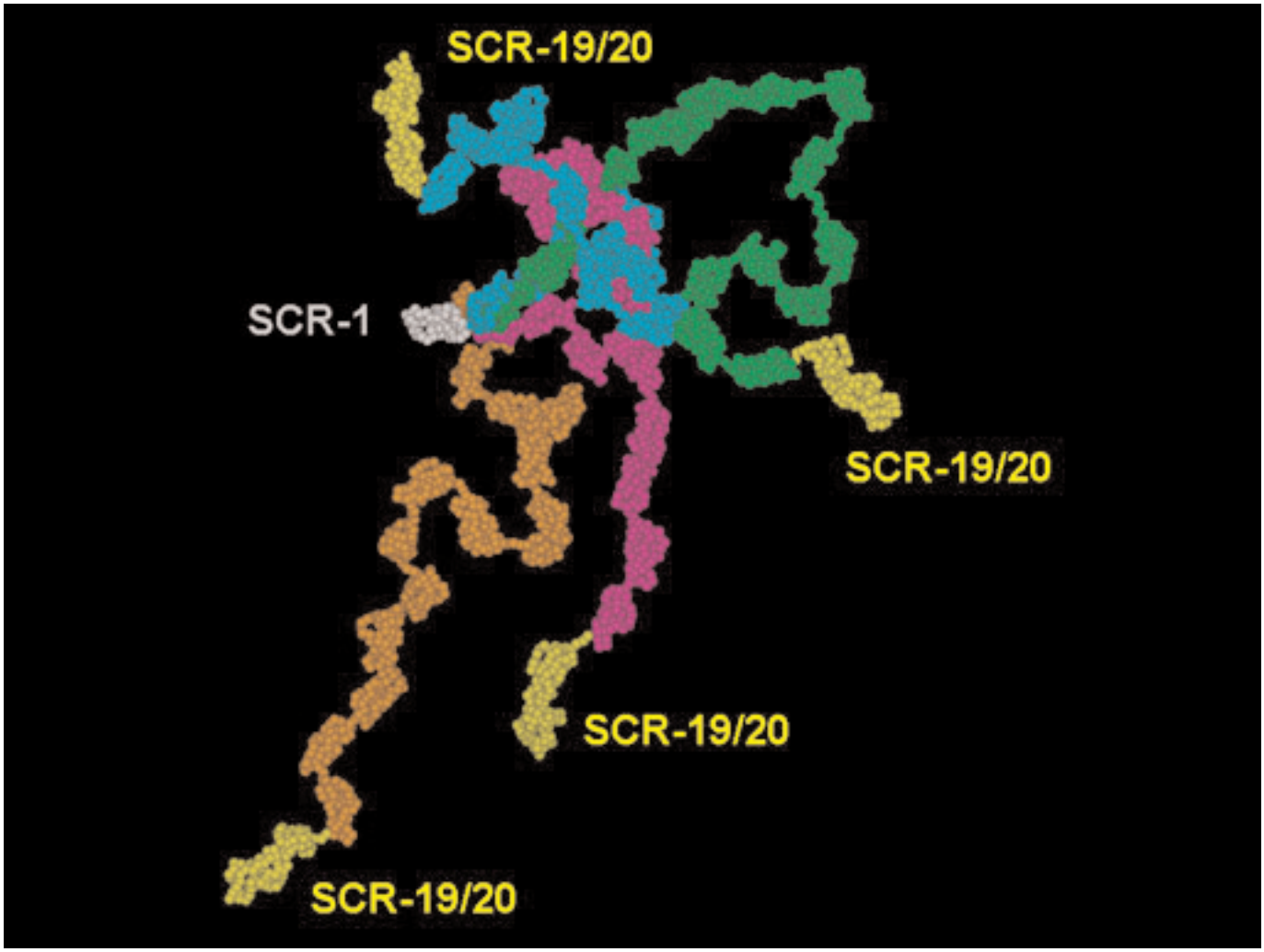

Figure 4: Positional summary of the four best-fit models for FH, all four of which are overlaid on SCR-1 (grey). SCR-19 and SCR-20 are highlighted in yellow.

into closer proximity. It is not clear at present why bent-back structures should occur, even though the known inter-domain orientations between SCR domains are variable. The most likely reason is the existence of five unusually-long linkers of lengths up to 8 residues between SCR-10/11, SCR-11/12, SCR-12/13, SCR-13/14 and SCR-18/19. These are mostly at the centre of the FH structure and may impart a higher degree of local flexibility to the FH structure to result overall in a structure that on average is not as extended as it could be. The automated rigid-body modelling strategy itself, which is based on 19 variable linkers here, as opposed to the single one used as a pair in antibody modelling, is seen to be sufficiently versatile to be successfully applied to this system.

The biological utility of this medium resolution modelling was demonstrated by the use of the four best-fit FH scattering models to interpret the significance of 12 missense mutations found in FH that result in haemolytic uraemic syndrome, a disease involved with renal disorders and kidney failure [5]. Ten mutations occur in SCR-19 and SCR-20 (yellow domains in Figure 4). Eight of nine mutations in SCR-20 are shown in Figure 5a (two involve the same residue), while the tenth one occurs at D1101 in SCR-19. The homology model for SCR-20 [1] showed that the majority of these nine mutations are immediately adjacent to basic

conserved residues in SCR-20, that in turn could be correlated with a predicted heparin binding-site in SCR-20, for which a model could be built. Since the amino acid substitutions occur in both SCR-20 and SCR-19, the interdomain orientation between these was of interest. Despite the possibility that the three-residue linker connecting the two domains can vary in its conformation, Figure 5b shows that the range of this variation is restricted in the four best-fit models. The substitution at D1101 in SCR-19 (four yellow circles) is close to the predicted heparin binding-site in SCR-20 and may affect the linker conformation. The scattering models therefore show that SCR-19 may contribute towards the formation of a heparin binding site in FH. We conclude that the medium resolution models generated by scattering modelling have utility for the interpretation of biological function.

Acknowledgements

We thank the Wellcome Trust and the BBSRC for their support, and Dr T. H. J. Goodship and Dr R. B. Sim for many useful discussions.

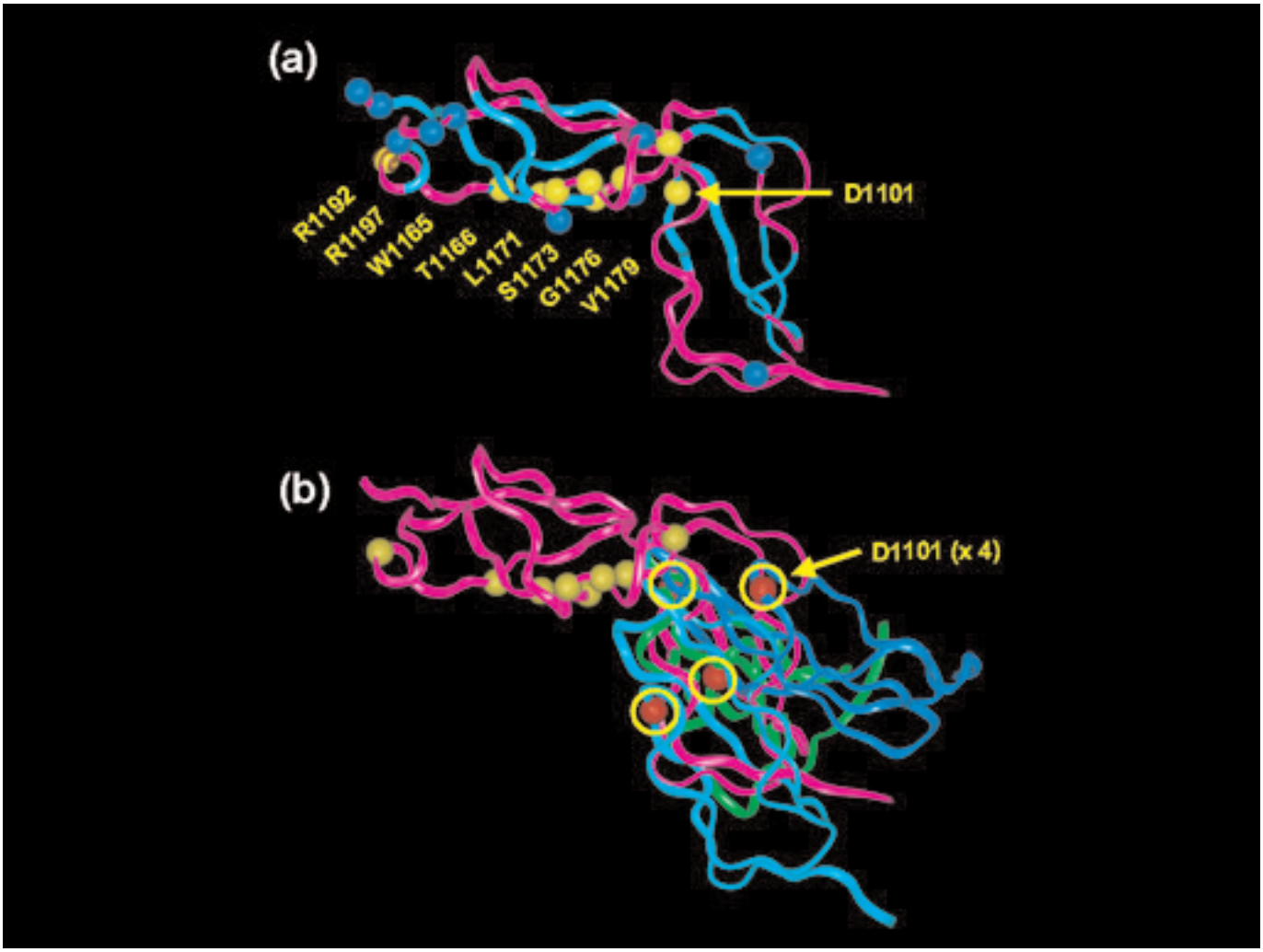

Figure 5: Use of the FH scattering models to provide a molecular explanation for haemolytic uraemic syndrome. (a) The single residue substitution in SCR-19 and nine in SCR-20 are shown as nine yellow spheres in one of the best-fit models (purple ribbon), together with the basic residues denoted as blue spheres. The b-strands are shown as blue ribbons. (b) SCR-19 and SCR-20 in the four best-fit models are shown superimposed on SCR-20, with the four positions of D1101 in SCR-19 denoted by yellow circles. The other three models are shown as blue and green ribbons.

References

[1] Aslam, M. & Perkins, S. J. (2001) Folded-back solution structure of monomeric Factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol*. **309**, 1117-1138.

[2] Perkins, S. J., Ashton, A. W., Boehm, M. K. & Chamberlain, D. C. (1998). Molecular s tructures from low angle X-ray and neutron scattering studies. *Int. J. Biol. Macromol*. **22**, 1-16.

[3] Perkins, S. J. (2001). Applications of highly constrained molecular modelling scattering curve fits to biologically important proteins. *Fibre Diffraction Review*, **9**, 51-58.

[4] Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999) The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: A study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol*. **286**, 1421-1447.

[5] Perkins, S. J. & Goodship, T. H. J. (2002) Molecular modelling of mutations in the C-terminal domains of factor H of human complement: a new insight into haemolytic uraemic syndrome *J. Mol. Biol*. **316**, 217-224.

[6] Perkins, S. J., Nealis, A. S. & Sim, R. B. (1991) Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry*, **30**, 2847-2857.